

VOLUME 64
NUMBER 6

WHOLE No. 312
1950

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

A Scale for Measuring Teacher- Pupil Attitudes and Teacher- Pupil Rapport

By

CARROLL H. LEEDS

*Furman University
Greenville, S.C.*

Accepted for publication, October 8, 1949

Price \$1.00

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

Volume 50, No. 1
1950

Volume 50, No. 1
1950

Psychological Monographs

General and Applied

Published by the American Psychological Association
540 North Dearborn Street, Chicago 10, Illinois

Subscription price, \$4.00 per year in advance

COPYRIGHT, 1950, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION

A Scale for Measuring Teacher Pupil Attitudes and Teacher Pupil Response

CARROLL L. FENN
University of Illinois at Chicago

Chicago, Illinois
U.S.A.

Published by the American Psychological Association

THE AMERICAN PSYCHOLOGICAL ASSOCIATION
540 NORTH DEARBORN STREET, CHICAGO 10, ILLINOIS

TABLE OF CONTENTS

| | |
|--|----|
| I. THE PROBLEM AND ITS SETTING | 1 |
| II. THE CONSTRUCTION, ADMINISTRATION, AND SCORING OF THE TEACHER-PUPIL INVENTORY | 3 |
| A. CONSTRUCTION OF THE ORIGINAL INVENTORY | 3 |
| 1. Nature of the Measuring Instrument | 3 |
| 2. Construction of the Items | 3 |
| 3. Classification of the Items | 3 |
| 4. Two Forms of the Inventory | 4 |
| B. ADMINISTRATION OF THE ORIGINAL INVENTORY | 5 |
| 1. Empirical Basis for Standardization | 5 |
| 2. Nature of the Teacher Population | 5 |
| 3. Personal and Descriptive Data | 6 |
| 4. Explanation and Directions for Completing the Inventory | 7 |
| 5. Distribution of the Inventories | 7 |
| C. COMPARISON OF RESPONSES AND SELECTION OF ITEMS | 8 |
| 1. Comparison of Responses to the Items | 8 |
| 2. Selection of Items for the Final Inventory | 10 |
| D. THE ORIGINAL SCORING SYSTEM FOR THE INVENTORY | 11 |
| 1. Procedure Used in Calculation of Weights for Item Responses.. | 11 |
| E. THE SIMPLIFIED SCORING SYSTEM FOR THE INVENTORY | 11 |
| F. COMPARISON OF INVENTORY SCORES OF THE SUPERIOR AND INFERIOR GROUPS OF TEACHERS | 11 |
| 1. Distribution of Inventory Scores for the Superior and Inferior Groups of Teachers | 12 |
| III. THE VALIDITY AND RELIABILITY OF THE TEACHER-PUPIL INVENTORY | 13 |
| A. DETERMINING THE VALIDITY OF THE INVENTORY | 13 |
| 1. Plan of Procedure | 13 |
| 2. Administration of the Revised Inventory | 13 |
| a. The nature of the teacher population | 13 |
| b. Mechanics of administration | 14 |

| | |
|---|----|
| 3. Obtaining the Ratings of the Teachers | 14 |
| a. Ratings by principal | 14 |
| b. Classroom observation | 15 |
| c. Attitudes of pupils | 16 |
| d. The reliability of the ratings | 17 |
| (1) Use of split-half technique | 18 |
| (2) Reliability of the scale used by the principal | 18 |
| (3) Reliability of the scale used by the writer | 18 |
| (4) Reliability of the questionnaire used by the pupils.. | 18 |
| 4. Comparisons Among Ratings and Inventory Scores | 19 |
| a. The distribution of Inventory scores for the unselected group of teachers | 19 |
| b. Correlations between Inventory scores and rating scale scores | 20 |
| 5. An Additional Check on the Validity of the Items | 21 |
| B. DETERMINING THE RELIABILITY OF THE INVENTORY | 22 |
| IV. SUMMARY AND CONCLUSIONS | 23 |
| BIBLIOGRAPHY | 24 |

A SCALE FOR MEASURING TEACHER-PUPIL ATTITUDES AND TEACHER-PUPIL RAPPORT¹

I. THE PROBLEM AND ITS SETTING

IN VIEW OF the vast amount of educational research carried on in recent years, it is rather surprising that one very important problem has almost escaped the critical attention of investigators. This is the problem of relationship between teacher and pupil, or the more specific phase of this relationship as manifested in the personal interaction of teacher and pupil in the classroom. There are far too many classrooms today in which the social atmosphere is little above the level of barbarism with reference to teacher-pupil interaction. Too often, pupils are treated with little respect, discipline problems rob the teacher of energy needed elsewhere, and classrooms are governed by individuals who are seriously maladjusted in personality. Although recent years undoubtedly have witnessed a growing interest in this problem, research workers have contributed little toward its solution, presumably because of the difficulty a scientific approach into the field would present. Modern educators and school administrators are, in most cases, fully cognizant of the problem, but a general feeling of helplessness with reference to its solution has prevented an energetic effort to remedy the undesirable situation. Granted the importance of the problem, it would seem that an instrument designed to measure certain aspects of

teacher-pupil interaction would prove of great value in the hands of educators and school administrators.

The purpose of this study is to explore the possibility of constructing a measuring instrument that will aid in the differentiation of teachers who have or will have rapport with pupils, from those who do not or will not have such rapport. It is postulated, of course, that rapport between teacher and pupil constitutes one of the many factors essential to teaching success. Although only one of many such factors, it is assumed further that it is one of the most important. Rapport between teacher and pupil refers to a state of harmonious relationship characterized by mutual affection, sympathy, understanding, and co-operative behavior. Definitely involved is the teacher's personality which, on assumption, either facilitates or checks the establishing of satisfactory rapport with pupils. Numerous studies have indicated the importance of personality to success in teaching. However, the present study is not directly concerned with prognosis of success in teaching, and the personality factor, affecting teacher-pupil rapport as it does, is assumed merely as *one very important* element in both the vocational life of the teacher and the learning career of the pupil.

As attitudes and feelings probably represent the core of human relationships in all of their manifestations and as the area covered by teacher-pupil relationships is quite large, consideration, in general, will be restricted to these re-

¹ The major part of a Ph.D. investigation conducted at the University of Minnesota under the direction of Dr. Walter W. Cook, Professor of Education. To Dr. W. W. Cook and Dr. T. W. McConnell the writer gratefully acknowledges his indebtedness for a challenging problem and the inspiration and helpful ideas so essential to an effective program of attack.

lations as exemplified in the attitudes and feelings each group holds toward the other. It will be assumed that a teacher's attitude toward pupils and toward children in general is an index to the rapport he has or will have with them.

Rapport between two people involves relationship in two directions. Not only, then, must we consider the teacher's attitude toward pupils, but also how the pupils feel toward the teacher whose attitude has been determined. In the present study the possibility is explored of validating the instrument for the

measurement of a teacher's attitude by determining the attitude of the pupils toward that teacher. Thus, the reaction of the pupils toward each particular teacher in the study represents one of the criteria against which the present instrument is validated. Presumably, a relationship between teacher and pupil in which the teacher holds an unsatisfactory attitude would also be characterized by antagonism or resentment on the part of the pupil. At least, this represents a possibility which this study attempts to explore.

II. THE CONSTRUCTION, ADMINISTRATION, AND SCORING OF THE TEACHER-PUPIL INVENTORY

A. CONSTRUCTION OF THE ORIGINAL INVENTORY

1. *Nature of the Measuring Instrument*

THE THURSTONE TECHNIQUE of scale construction (20) was rejected principally because its rigid demand for linearity along a degree-of-favorableness continuum would seem to present an undesirable restriction and limitation. Furthermore, the use of judges or experts (as in the Thurstone technique) was not contemplated in the present study.

Although the Likert technique (11) was not adopted in its entirety, the type of instrument finally decided upon is similar in some respects to that of Likert's. The multiple-choice mode of response to opinion statements (each item thus approximating a scale in itself) was employed. The assigning of numerical values, for scoring purposes, to each of the several responses per item also represents a point of similarity, although the method of weighting these responses differs considerably from that used in the Likert technique.

The instrument was to be known as the *Teacher-Pupil Inventory*.

2. *Construction of the Items*

Opinion statements were chosen in preference to concrete situations with multiple-choice reactions. It was felt that a relatively large number of opinion statements, touching upon a variety of topics, would tap more adequately a teacher's attitude toward pupils than would fewer and less flexible items of the other type. In making this choice, the investigator was fully aware of the discrepancy which so often appears be-

tween *opinion* and *overt action*. Yet, in the appraisal of *attitude*, opinion is perhaps as valid an indicator as is potential overt behavior.

The subject was to respond to each item by encircling a number representing the degree of his acceptance as follows: (1) strongly agree, (2) agree, (3) undecided or uncertain, (4) disagree, (5) strongly disagree. The following example illustrates a "strongly agree" reaction:

112. Immediate obedience is not always desirable.

① 2 3 4 5

An extensive survey of the literature was made to obtain a wealth of ideas concerning all phases of teacher-pupil relationship, both in and out of the classroom, to be used as source material for the building of items. These ideas, coupled with those from the writer's own experience, provided the basis for the compilation of a list of 378 opinion statements relating to the reaction of teachers to children and pupils and their behavior. Items expressing opinions concerning children in general were included along with those relating directly to pupils, as it was felt that a teacher's reaction to *children as such* is of significance in her attitude toward pupils. Items were constructed with children of elementary school age primarily in mind. It was believed that the Inventory eventually would prove most useful at this level.

3. *Classification of the Items*

On the assumption that the employment of some system of classification in item construction would contribute to a more adequate sampling of teacher

attitude and behavior in relation to pupils, the opinion statements were classified under five categories which are presented here with brief explanatory comments and illustrative items:

(1) *Moral status*. This phrase is meant to cover those adult-child relations essentially characterized by adult recognition or non-recognition of child or pupil as an individual with his own inherent rights. Placed under this category were items that related to the status of children in the eyes of adults, especially as concerns their adherence to adult-imposed standards, moral and otherwise.

Example: Children "should be seen and not heard."

(2) *Discipline*. Under this caption were included those statements relating to problems of conduct in the classroom and elsewhere, and the methods employed in dealing with them.

Example: Pupils found writing notes should be severely punished.

(3) *Child knowledge*. Here were placed those statements relating to principles of child development and behavior ordinarily included in such subject fields as educational psychology, child psychology, and mental hygiene. More specifically, the content of these items related to the topics of ability, achievement, learning, motivation, and personality development.

Example: The boastful child is usually over-confident of his ability.

(4) *Educational principles*. Under this category were placed items pertaining to educational principles and practices not directly relating to child development and behavior. Problems in education concerning method, philosophy, curriculum, and administration are touched upon here.

Example: Pupils should be required to do more studying at home.

(5) *Personal reactions of teacher*. Included under this category were statements directly relating to the teacher's personal disposition—her likes and dislikes, sources of irritation, etc.

Example: Without children life would be dull.

Consideration of the above system of classification will reveal a rough attempt to identify psychological factors in personality structure and function, as well as to make a differentiation relative to the ideational content of items. It was believed, for example, that the categories "Moral status" and "Personal reactions of

teacher" related to responses that involved more of the affective nature and less of the cognitive, than did the other categories. The feeling component in these areas would seem to be stronger than the intellectual element. At least, the subject-matter that would make up the content of items under these two categories was less objective and less well established as ideational principles. However, the composition of psychological reactions to statements of opinion, as expressed in terms of affective and intellectual elements, is still essentially a matter of conjecture.

Added to this uncertainty is the problem of overlapping among the categories. As the classification of statements proceeded, the difficulty of the process became more and more apparent, with an increasing realization that a number of the items should be classified under two or more categories because of a very evident relevance to each of several topics. The following item, for example, was classified under the categories of "Moral status," "Discipline," and "Educational principles":

Unquestioning obedience in a child is not desirable.

And, who is to say that this item, with possibly equal justification, could not have been placed under the two remaining categories of "Child knowledge" and "Personal reactions of teacher?"

Such considerations as the above made the whole problem of item classification very difficult to handle, especially on a logical and an a

TABLE I
NUMBER OF ITEMS IN FINAL INVENTORY OF
164 ITEMS CLASSIFIED UNDER EACH OF
FIVE CATEGORIES

| Category | No. of Items |
|-------------------------------|--------------|
| Moral status | 110 |
| Discipline | 43 |
| Child knowledge | 43 |
| Educational principles | 50 |
| Personal reactions of teacher | 43 |

priori basis. The classification process was completed, however, as it was felt that it would lead to a more substantial body of inventory content. Table I indicates the number of items in the final inventory of 164 items that were classified under each of the five categories. A number of the items were placed in more than one division.

4. Two Forms of the Inventory

To facilitate the selection of the best-stated and most discriminating items,

two forms (A and B) of the *Teacher-Pupil Inventory* were constructed. Form B was made up of a second list of 378 items, or opinion statements, each of which (with the exception of 8 or 10 items) was stated in a different manner (usually in reverse form) from a correspondingly numbered related item in the first list (Form A). Samples of several items from Form A with correlative items from Form B are given below.

1. Most children are obedient. (Form A)
1. Most children tend to be disobedient. (Form B)
2. Pupils expect too much help from the teacher in getting their lessons. (Form A)
2. Teachers should give pupils more direct help with their lessons. (Form B)
3. The majority of children take their responsibilities seriously. (Form A)
3. Most children are irresponsible. (Form B)

In order to make Form A and Form B as equivalent as possible, approximately 50 per cent of the items in each form were so worded that agreement with these statements corresponded to the same end of the attitude continuum as did disagreement with the other 50 per cent of the items. These two kinds of statements were distributed at random throughout each form. Items representing the five categorical topics were also placed at random within each form. The number of items in each form (378) was made amply large to provide for the elimination of poor items.

B. ADMINISTRATION OF THE ORIGINAL INVENTORY

1. *Empirical Basis for Standardization*

Instead of using judges or experts to rate the items as to validity and quality, to establish a basis for standardization, and to devise a system of scoring, it was the plan of the investigation to arrive at these ends by *empirical* means. That is,

the plan was to administer the Inventory to two groups of teachers (one group characterized as maintaining good relations with pupils, the other group as maintaining poor relations) to determine the *differences*, if any, between the groups in their reaction to the items. In other words, instead of logically determining in advance how certain teachers *ought* to react to the items and then seeing if they measure up to this standard, the method of empiricism involves the experimental discovery of how various teachers actually *do* react to the items, which finding then serves as a basis for the interpretation of all subsequent reactions.

The criterion used in the original selection of the two groups of teachers was the local school administrator's judgment concerning the teacher-pupil relationship status of his individual faculty members. In employing this criterion, the investigator fully realized that administrators often are not as aware as they should be of the actual nature of the relationship existing between the pupils and individual members of the teaching staff.

2. *Nature of the Teacher Population*

The teacher population to which the original *Teacher-Pupil Inventory* was administered comprised two groups of teachers, 100 in each group, differentiated on the basis of the local school administrator's judgment concerning the individual teacher's ability to maintain harmonious relations with pupils. This was the only point of differentiation that was controlled in the selection of the two groups. Each group was heterogeneous in nature as there was no attempt made to control such factors as age, sex, marital and parental status, nationality,

grade level, subject taught, training, and experience. Little consideration was given these factors at this stage of the investigation, as the essential focus of attention here was the *difference* in reaction to items of the Inventory between the two groups of teachers who had been differentiated solely on the basis of relationship with pupils. The writer was aware of the possible influence of the

the investigator's judgment they were, in general, educational institutions average in philosophy and practice, probably tending toward the traditional and conservative type of school.

Table 2 indicates the extent to which different-sized communities were represented in making up the two groups of 100 teachers each. The representation, in each group, of the different school levels is also indicated. No effort was made to equate the two groups relative to the size of school system or to the grade level represented. Teachers at each extreme of the teacher-pupil relationship continuum were used irre-

TABLE 2
NUMBER OF SCHOOLS AND TEACHERS REPRESENTING DIFFERENT-SIZED COMMUNITIES AND
DIFFERENT SCHOOL LEVELS FROM WHICH TWO GROUPS OF 100 TEACHERS
EACH WERE DRAWN

| Size of Community | Superior Group | | | Inferior Group | | |
|---|-------------------|-----|-----------------|-------------------|-----|-----------------|
| | No. of Schools | | No. of Teachers | No. of Schools | | No. of Teachers |
| Cities with population of 100,000 and over | Sr. High | 2 | 3 | Sr. High | 1 | 4 |
| | Jr. High | 1 | 1 | Jr. High | 1 | 2 |
| | Elem. | 9 | 14 | Elem. | 8 | 16 |
| Cities with population between 1,000 and 100,000 | Sr. High | 9 | 21 | Sr. High | 8 | 20 |
| | Jr. High | 5 | 10 | Jr. High | 4 | 9 |
| | Elem. | 17 | 30 | Elem. | 18 | 33 |
| County schools and towns with population of 1,000 and under | High } Elem. } | 14* | 9 12 | High } Elem. } | 11* | 9 7 |
| | — | | — | — | | — |
| | 57 | | 100 | 51 | | 100 |

* Both elementary and secondary grades housed in one building.

other factors upon teachers' attitudes, but consideration of this problem was reserved until a later time.

An effort was made to contact teachers from schools of varying size and description so as to include individuals with varied teaching environments. However, the study was limited to teachers in the public schools at the elementary and high school levels.

The schools included were located in Pennsylvania and Ohio and ranged in size from one-room county schools to the elementary and high schools of one of our larger cities. The schools represented farming communities, industrial and mining regions, mercantile areas, and wealthy residential districts. None of the schools was extremely progressive or ultraconservative. In

spective of where they were found. Reference to Table 2, however, will reveal a close similarity in composition of the two groups.

3. Personal and Descriptive Data

By means of a Form for Personal and Descriptive Data which accompanied the Inventory, a record was kept, for each teacher, of the factors of age, sex, marital and parental status, nationality, grade level, subject taught, training, and experience. The teacher was also asked (a) to indicate whether or not he had had a course in mental hygiene, and (b) to rate himself relative to his liking for teaching. The name of the teacher was not requested. It was felt that anonymous

responses would approximate more closely actual attitudes.

4. *Explanation and Directions for Completing the Inventory*

Explanation and instructions relative to completing the Inventory were as follows:

Explanation and Directions

There is considerable disagreement in educational circles about a variety of topics involving the relationships of teachers and pupils. These topics are still largely matters of opinion with little reinforcement from scientific evidence. It is the purpose of the Inventory to sample existing opinion in the area of teacher-pupil relationships. There are no right or wrong answers to the items of the Inventory. *Your own individual feeling* about the matter is what is wanted. A *frank and sincere* response to each item is obviously necessary if this study is to serve its purpose. Your responses will have no effect whatever upon your personal or professional standing. *Your name is not requested.*

Directions: After reading the item carefully, indicate your response by encircling the number which most nearly represents your reaction to the topic in question. Think in terms of general situations rather than of individual cases. *Please respond to every item.* The scale is as follows:

1. Strongly agree
2. Agree
3. Undecided or uncertain
4. Disagree
5. Strongly disagree

5. *Distribution of the Inventories*

Before the teachers were approached, permission to carry on the investigation in the particular school system and building was obtained from the superintendent and principal, respectively, and their cooperation was assured. The purpose and methodology of the study were carefully explained to them. The school official feeling most qualified (in most cases this was the principal) was asked to list, if possible, several teachers on his staff whom he considered quite superior in their relationships with pupils and several at the opposite extreme, whom

he considered quite inferior in this respect. In making this selection, he was to consider the following criteria with reference to each teacher:

- (1) Ability to win the affection of his pupils.
- (2) Fondness for, and understanding of children.
- (3) Ability to maintain a desirable form of discipline.

Although the characteristics above are not necessarily correlative, the school official generally experienced little difficulty in making the selection desired. It was made clear to him that, among a number of factors that go to make up the professional requirements of a teacher, relationship with pupils was the only one that he was to consider in his selection. Such factors as intelligence, teaching methods, and knowledge of subject-matter were not to be used as differentiating criteria.

The investigator presented a copy of Form A of the original Inventory to each of the 200 teachers in person.¹ Each was asked if he would cooperate in a research project by completing the Inventory sometime within the next several weeks, at the end of which time the writer would return for it. Each teacher was assured that no one but the writer would have access to the completed Inventory. It was made clear that the project was the writer's only, and that the school system or administration had had no part in the instigation of the study. Although there were a few teachers who did not wish to participate, cooperation, for the most part, was excellent. For purposes of identification, a record was kept of the number of the copy that was given to each teacher.²

¹ Although consuming much time and effort, this individual, personal contact was considered essential to the success of the investigation.

² This number had been placed on the Inventory quite inconspicuously (immediately following the form designation; e.g., Form A6).

Thus, from none to five or six teachers at each extreme (depending, among other factors, upon the size of the teaching staff) were located in each of sixty-seven buildings that were visited in fourteen school systems. In this manner there were procured 100 subjects in each of two extreme groups—one group composed of teachers who were considerably above average in their ability to get along with pupils (to be known as the "superior" group), and the other group consisting of teachers considerably below average in this ability (to be known as the "inferior" group).

After a month or so had elapsed, the writer revisited the sixty-seven schools and again saw, individually, each of the 200 teachers. At this time Form A of the Inventory was taken up and a copy of Form B, with a return envelope, was left with each teacher. For purposes of identification care was taken that the number on the copy of Form B given to any one individual corresponded with that on his copy of Form A.

C. COMPARISON OF RESPONSES AND SELECTION OF ITEMS

As soon as the Inventories were completed and returned, the response of each of the 200 teachers to each of the 378 items in each of the two forms of the Inventory was carefully recorded. The

purpose of this work was to *compare* the responses of the two groups of teachers to each of the items in order to determine possible *differences* in reaction. Items that did not differentiate the groups could then be discarded and a scoring system devised for the remaining items on the basis of the obtained differential responses.

1. Comparison of Responses to the Items

The frequency of each of the five responses (1, 2, 3, 4, 5) for each item was obtained, separately, for each group. As there were 100 individuals in each group, the frequencies also took the form of percentages. As illustrated in Table 3, a table showing the pattern of response was constructed for each of the 756 items (Forms A and B).

The response pattern in Table 3 is interpreted as follows: 174, or 87 per cent, of the 200 teachers are in agreement with this item (including both the "strongly agree" and "agree" responses). Thirty-four per cent of the superior group "strongly agree"; 18 per cent of the inferior group "strongly agree"; 3 per cent of the superior group "disagree," whereas 13 per cent of the inferior group "disagree"; the same percentage of each group are "undecided" (4 per cent); the same percentage "strongly disagree" (1 per cent). It is of interest to note that a larger percentage of the inferior group (64 per cent) than of the superior group (58 per cent) "agree" to the item. This apparent discrepancy is probably explained by the fact that a relatively larger proportion of the superior group is found in the "strongly agree" column. This ex-

TABLE 3
DISTRIBUTION OF RESPONSES OF THE SUPERIOR AND INFERIOR GROUPS OF
TEACHERS TO ITEM 1 (FORM A)

| 1. Most children are obedient. | | | | | |
|--------------------------------|----------------|-------|-----------|----------|-------------------|
| | Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree |
| Group | 1 | 2 | 3 | 4 | 5 |
| Superior | 34 | 58 | 4 | 3 | 1 |
| Inferior | 18 | 64 | 4 | 13 | 1 |
| Difference | 16 | -6 | 0 | -10 | 0 |

TABLE 4
DISTRIBUTION OF RESPONSES OF THE SUPERIOR AND INFERIOR GROUPS OF
TEACHERS TO ITEM 243 (FORM A)

| 243. If the teacher laughs with the pupils in amusing classroom situations, the class tends to get out of control. | | | | | |
|--|----------------|-------|-----------|----------|-------------------|
| | Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree |
| Group | 1 | 2 | 3 | 4 | 5 |
| Superior | 1 | 0 | 1 | 55 | 43 |
| Inferior | 1 | 4 | 6 | 67 | 22 |
| Difference | 0 | -4 | -5 | -12 | 21 |

planation became more and more plausible as the reactions to other items were studied. Consider the response pattern in Table 4 as a second example. Over 90 per cent of the teachers do not agree with this item. The most differentiating response between the groups is No. 5 ("strongly disagree"). It seems reasonable to

other words, one might characterize their reaction as one of *nominal disagreement*, or as representing the most widely accepted opinion, whereas the reaction of the superior teachers is strong enough to indicate an attitude of sufficient potency to effectively influence their daily professional work. With a number of items, in

TABLE 5
DISTRIBUTION OF RESPONSES OF THE SUPERIOR AND INFERIOR GROUPS OF
TEACHERS TO ITEM 290 (FORM A)

| 290. Most pupils don't appreciate what a teacher does for them. | | | | | |
|---|----------------|-------|-----------|----------|-------------------|
| | Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree |
| Group | 1 | 2 | 3 | 4 | 5 |
| Superior | 0 | 38 | 16 | 39 | 7 |
| Inferior | 7 | 49 | 15 | 29 | 0 |
| Difference | -7 | -11 | 1 | 10 | 7 |

interpret the relatively high percentage (67 per cent) of the inferior group found under Response 4 ("disagree") as a general disagreement on the part of the teachers of this group concerning the item, but as an apparent unwillingness to go to the extreme in their response. In

this manner, the strength of differentiation lay in the difference in response of the groups relative to the two degrees of reaction at either end of the attitude continuum.

The responses indicated in Table 5 assume a different pattern. The responses for this item

TABLE 6
DISTRIBUTION OF RESPONSES OF THE SUPERIOR AND INFERIOR GROUPS OF
TEACHERS TO ITEM 164 (FORM B)

| 164. As a rule pupils do not have much use for teachers. | | | | | |
|--|----------------|-------|-----------|----------|-------------------|
| | Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree |
| Group | 1 | 2 | 3 | 4 | 5 |
| Superior | 1 | 4 | 3 | 67 | 25 |
| Inferior | 1 | 7 | 15 | 61 | 16 |
| Difference | 0 | -3 | -12 | 6 | 9 |

assume a rather striking pattern of symmetry. A larger proportion of the inferior group, however, are in agreement with the item. It is of interest to note that the same number (7) of inferior teachers "strongly agreed" to the item as superior teachers who "strongly disagreed." There is also a remarkably close agreement between the groups relative to Response No. 3 ("undecided"). This similarity of response in the "undecided" category is characteristic of a number of the items.

There are other items, however, with which considerable differentiation between the groups occurs in the "undecided" response. Table 6 illustrates an item in which most of the differentiation occurs in this category. Although with this item 84.5 per cent of the 200 teachers are in disagreement, *undecidedness* seems to characterize the inferior group more than it does the other group.

2. Selection of Items for the Final Inventory

Only 164 of the original 756 items were chosen for use in the final Inventory, the remaining items being rejected. The criteria employed in the selection and rejection of items were as follows:

- (1) Was the item adequate in differentiating the two groups of teachers?
- (2) Was the item ambiguous in meaning, lacking in clearness, or poorly stated?
- (3) Did the content of an item duplicate that of another item that had been selected?
- (4) Did the item show a response pattern that was difficult to interpret?

The statistic, chi-square, was employed as an objective measure of the degree of differentiation between two response patterns. The following formula (7, pp. 385-386) was used to calculate the contribution to chi-square of any given pair of responses:

$$\frac{1}{a + a'} (aN' - a'N)^2 \text{ where,}$$

a and a' represent the obtained column entries and N and N' represent the corresponding row totals.

As a valid and practical measure of the degree

of differentiation between two response patterns, chi-square was found to have its limitations. Large chi-square values were sometimes the result of differences in response that were found in two categories only (of which one was often the "undecided" response) and of differences in response that were irregularly distributed throughout the attitude continuum, thereby making interpretation difficult. A few items with relatively high chi-square values were, consequently, rejected for use in the final Inventory. On the other hand, there were a few items with relatively low chi-square values which were selected for use because of their symmetrical response patterns or because their inclusion in the Inventory would lead to a richer and more diversified body of content for the instrument.

The value of chi-square proved to be an effective criterion for item selection provided the remaining three criteria were satisfied. Generally speaking, an item (even though possessing a high chi-square value) was discarded if it did not meet the demands of all three remaining criteria.

Of the 164 items chosen for use in the final Inventory:

- 79 (48 per cent) met the probability level of .05 ($\chi^2 = 9.488$)
- 112 (68 per cent) met the probability level of .10 ($\chi^2 = 7.779$)
- 134 (82 per cent) met the probability level of .20 ($\chi^2 = 5.989$)
- 146 (89 per cent) met the probability level of .30 ($\chi^2 = 4.878$)
- 157 (96 per cent) met the probability level of .50 ($\chi^2 = 3.357$)

It was in the original plan of the investigation to so construct the *Teacher-Pupil Inventory* that, in its final state, it would contain from 150 to 175 items in each of two forms. As analysis of items progressed, however, it became increasingly apparent that the final instrument would be more satisfactory if it consisted of one form with a more select group of items than of two forms with inferior items. The revised Inventory, then, is-

sued in one form containing 164 of the most select items.

D. THE ORIGINAL SCORING SYSTEM FOR THE INVENTORY

The simplified scoring system of Likert (use of consecutive numerical values 1, 2, 3, 4, 5, assigned to the several responses) had been contemplated for use with the present instrument. When responses to the items were analyzed, however, the idea of using this technique was abandoned. It became increasingly evident that, in the great majority of cases, the responses for each item were not normally distributed. Rough inspection revealed tendencies toward distributions bi-modal in character, or with a piling up of responses at one or the other end of the attitude continuum. Furthermore, no assumption was to be made relative to uniformly—or non-uniformly—spaced “steps” along the attitude continuum.

1. *Procedure Used in Calculation of Weights for Item Responses*

Consideration of such factors as the above, coupled with the desire to make the investigation as empirical as possible, led to the adoption of a procedure for weighting the several responses for each item in accordance with the *differentiation* empirically established between the two groups of teachers studied. This was the method employed by Strong (17) in weighting the responses to the items of his Vocational Interest Inventory. Strong's formula, contributed by T. L. Kelley (17, pp. 603-615), was applied to the construction of the scoring key for the *Teacher-Pupil Inventory*. The range of weights used extended from plus 4 to

minus 4. Weights beyond these limits were valued at plus 4 and minus 4, respectively. For each of the five responses associated with each of the 164 items in the revised Inventory the formula was used to compute the weight to be assigned to that response. The value of *W* obtained by use of the formula is multiplied by 100, and the nearest whole number is used.

E. THE SIMPLIFIED SCORING SYSTEM FOR THE INVENTORY

The system of scoring indicated above, with weights ranging from plus 4 to minus 4 assigned to the several responses, necessitates the rather cumbersome process of adding algebraically a fairly long list of positive and negative numbers. Later in the study it was found that results produced by a scoring scheme which allotted plus 1 to each circled response with a positive weight (disregarding all circled responses with negative and zero weights) correlated .973 with results attained by the longer method. This simpler scoring method was therefore later adopted.

F. COMPARISON OF INVENTORY SCORES OF THE SUPERIOR AND INFERIOR GROUPS OF TEACHERS

After the original scoring key had been established, it was used in the scoring of the 200 inventories received from the two groups of teachers. Forms A and B were now combined, and, in the scoring, consideration was limited to those 164 items which had been selected for the final form of the Inventory. Results obtained in comparing the scores of the superior and inferior groups of teachers

TABLE 7
RESULTS OBTAINED IN THE COMPARISON OF
INVENTORY SCORES OF THE SUPERIOR
AND THE INFERIOR GROUPS OF
TEACHERS

| Statistic | Superior Group | Inferior Group |
|--------------|-------------------|-------------------|
| N | 100 | 100 |
| Range | 324-(-99) | 213-(-297) |
| M | 131.0 | -32.0 |
| S.D. | 102.91 | 112.46 |
| σ_M | 10.19 | 11.18 |
| Mdn | 130.0 | -41.5 |
| σD_M | | 15.13 |
| D_M | | 163.01 |
| C.R. | | 10.8 |

are presented in Table 7. It may be seen that, despite some overlapping, there is a highly significant difference between

the highest score (324) was obtained by a Dean of Women in a high school and a score fifth from the top (295) was made by a junior high school teacher who also was serving in the capacity of visiting teacher. These were the only two individuals in the sample occupying such unique professional status.

1. Distribution of Inventory Scores for the Superior and Inferior Groups of Teachers

The statistics g_1 and g_2 , suggested by Fisher (15) as appropriate tests for the normality of a distribution, were calculated for each of the two sets of scores. The quantity g_1 is a measure of skewness; the closer its value to zero, the more symmetrical the distribution. The negative value of g_1 obtained for the superior group, as shown in Table 8, indicates a slight asymmetry with an excess of scores larger than the mean. The posi-

TABLE 8
VALUE AND SIGNIFICANCE OF g_1 AND g_2 FOR THE DISTRIBUTIONS OF INVENTORY SCORES MADE
BY THE SUPERIOR AND INFERIOR GROUPS OF TEACHERS

| Group | g_1 | t_{g_1} | Probability of t_{g_1} | g_2 | t_{g_2} | Probability of t_{g_2} |
|----------|--------|-----------|-----------------------------|--------|-----------|-----------------------------|
| Superior | -.0910 | .3789 | >.05 | -.8342 | 1.744 | >.05 |
| Inferior | .0927 | .3860 | >.05 | -.4472 | .9349 | >.05 |

the means of the two groups of teachers (critical ratio: 10.8). Although not shown in the table, 10 of the scores of the inferior group were above the mean of the superior group, and only 4 of the scores of the latter group were below the mean of the inferior group. Also, it is of interest to note that, in the superior group,

tive value of g_1 found for the inferior group indicates an excess of scores smaller than the mean. However, each of the g_1 values is too small to be significant.

The statistic g_2 is a measure of kurtosis. If its value is zero, normality is indicated so far as this measure is concerned. The negative values of g_2 obtained for the two distributions indicate flat-topped distribution curves. However, each of the g_2 values is too small to be significant.

III. THE VALIDITY AND RELIABILITY OF THE TEACHER-PUPIL INVENTORY

ALTHOUGH SUBJECTIVE in nature, the process of composing the items contributed to the validity of the instrument. In a more objective manner the administration of the Inventory to two groups of teachers at extreme ends of an attitude-toward-pupils continuum constituted a validating process in *establishing the validity of individual items*. Other criteria being satisfied, only items that served to *differentiate* the groups were selected for use in the final or revised instrument. The statistic, chi-square, served as the measure of the degree of differentiation.

It now remained to establish the validity of the Inventory as a whole.

A. DETERMINING THE VALIDITY OF THE INVENTORY

1. Plan of Procedure

The plan of determining the validity of the Inventory essentially involved (a) its administration to a new sample of 100 teachers, (b) the obtaining of ratings on these teachers with reference to their relationships with pupils, and (c) the correlation of the teachers' scores on the Inventory with the ratings. These ratings were to be obtained from each of three sources: (a) principals under whom the teachers were serving, (b) classroom observation on the part of the investigator, and (c) attitudes of pupils toward each individual teacher.

As a further check on the validity of the items of the Inventory, out of the new distribution of 100 teachers the item response patterns of the twenty with the highest Inventory scores and the response patterns of the twenty with

the lowest scores were to be compared, respectively, with the patterns of item response characterizing the original groups of superior and inferior teachers. Also, the responses of the twenty teachers with the highest scores were to be compared with the responses of the twenty with the lowest scores. The statistic, chi-square, was to be employed as the objective means of comparison.

2. Administration of the Revised Inventory

a. *The nature of the teacher population.* Inasmuch as the original sample of 200 teachers represented the two extremes in teacher-pupil relationship, it was deemed desirable to include in the new sample 100 teachers *unselected* with reference to their relationship with pupils. In designation, this group will be known as the "unselected" group.

The 100 teachers who were to serve as subjects were fourth-, fifth-, and sixth-grade teachers in three different school systems, two of which had been used in the earlier phase of the investigation. In these two systems, however, buildings different from those used previously were visited at this time. The one condition under which the 100 teachers were selected, aside from their status as public school employees, was that they have teaching contact with children in the fourth, fifth, or sixth grade at the time of testing. It was immaterial as to whether they were teaching one or more of these grades. Table 9 indicates the number of cities, schools, and teachers representing the different-sized communities from which the unselected group of 100 teachers was drawn.

TABLE 9
NUMBER OF CITIES, SCHOOLS, AND TEACHERS
REPRESENTING DIFFERENT-SIZED
COMMUNITIES FROM WHICH THE
UNSELECTED GROUP OF 100
TEACHERS WAS DRAWN

| Community Size | No. of Cities | No. of Schools | No. of Teachers |
|--|------------------|-------------------|--------------------|
| Cities with population of 100,000 and over | 1 | 5 | 46 |
| Cities with population between 1,000 and 100,000 | 2 | 13 | 54 |
| Total | 3 | 18 | 100 |

It was planned to use teachers of the elementary intermediate grades for two fundamental reasons: (a) Since the inception of the study, it had been the investigator's conviction that the prospective Inventory would prove of particular value at this important rung of the educational ladder. For where else than in the elementary grades is the matter of teacher-pupil relationship of such peculiar importance? Consequently, items were constructed with children of elementary school age primarily in mind. (b) It was believed that the attitudes of pupils toward their teachers (an important validating device in the study) could be obtained more effectively at the intermediate grade level. Pupils of grades four, five, and six would be of sufficient maturity to respond adequately to simple items in a questionnaire. Furthermore, it was felt that responses obtained from these pupils would be of a more ingenuous and less sophisticated nature than those that would be procured from pupils more advanced in educational achievement, who have far too often, become comfortably adapted to the conventional school situation.

b. Mechanics of administration. As in the previous visitation of schools, the initial step at this time was to explain the nature of the investigation to the administrative head of the particular school system and obtain the assurance of his cooperation in the study. After then interviewing the principals of the several buildings that were to be used and explaining the purpose and method of the investigation, the writer contacted all teachers at the fourth-, fifth-, and sixth-grade levels personally and individually.

Unless the teacher preferred not to participate in the study, a copy of the revised Inventory, together with a stamped return envelope, was left with him.¹

3. Obtaining the Ratings of the Teachers

Care was taken that no attempt was made to secure further data until all copies of the Inventory distributed to the teachers had been completed and returned. This precaution was considered essential, for any awareness on the part of the teacher that he was to be rated probably would have influenced his responses to the items in the Inventory. On the other hand, to avoid bias in the mind of the investigator in his ratings of teacher behavior, scoring of the Inventories was deferred until all data relative to rating of the teachers had been gathered.

a. Ratings by principal. A rating scale was devised on which a principal was to rate each of the teacher-subjects on his staff with reference to relationship with pupils. Ratings were to be obtained for each teacher on the following aspects of this relationship:

- (1) Disciplinary ability.
- (2) "Personnel" vs. "subject-matter" point of view.
- (3) Attitude toward children.
- (4) Understanding of pupil behavior problems.
- (5) Personality adjustment.
- (6) Attitude of pupils toward this teacher.

It would seem that the above items are important factors in the personal interaction between teacher and pupil. This personal interaction is the aspect of the whole teacher-pupil relationship pattern that is of particular interest in the present study. The topics mentioned

¹ The Inventory copies again had been numbered to make identification possible.

represent several basic areas of teacher-pupil relationship from which a number of the items of the Inventory were drawn.

The discipline that a teacher maintains in the classroom essentially is determined by the personal relationship that exists between him and the pupils. Whether it is the tyrannical teacher who "rules with an iron hand" and instills fear into his pupils or the supine teacher who is unable to "keep order" and whose pupils have no respect for him, the problem is principally one of affective and personal relationships.

The personal interaction between teacher and pupil probably is influenced by the teacher's own philosophy of education. At one extreme he may look upon the child merely as a passive receptacle for knowledge and dogmatically cling to "blanket" standards of achievement. Or, on the other hand, he may regard the mastery of subject-matter as secondary and actively strive to meet the varied needs of the child's developing personality.

Item number 3 in the Rating Scale, concerned with the teacher's attitude toward children, goes right to the heart of the problem with which the Inventory was designed to cope. Does the teacher look upon the child as a sinister and inferior being whose greatest virtue is to obey his elders? Or, at the other extreme, does the teacher naturally love children, and, sympathetically and understandingly, appreciate their individual problems of growing up?

Closely related to a teacher's attitude toward children is his understanding of child behavior problems. In his reactions toward the behavior of children does he indicate an understanding of, and an attempt to apply, the principles voiced by mental hygienists? For example, does he consider the behavior of the shy, withdrawing child as probably more serious than the aggressive transgression against authority? Does he actively seek to determine the causes behind misconduct, or is he satisfied with the consideration of symptoms only?

Of unquestioned influence in a teacher's personal interaction with pupils is the status of his own personality adjustment. It is a strong conviction of the writer that teachers with numerous disciplinary problems, teachers who are irritated easily by children's actions, and teachers who are often uneasy in the presence of children are beset with personality difficulties of their own which undermine their effectiveness in the schoolroom.

The attitude of the pupils toward a teacher is generally sensed by the principal. Especially is this true with a very popular teacher or with

one who is regarded as an enemy. As indicated elsewhere in this paper, the pupils' liking or disliking for a teacher is considered of basic significance in teacher-pupil relationship.

When all of the inventories had been returned, copies of the rating scale described above, together with mailing facilities for their return, were left with each of the twelve principals in the eighteen buildings visited in this part of the investigation. The discrepancy between the numbers of principals and buildings was due to the fact that, in two of the school systems, each principal had charge of two or three buildings.

The mean of the principals' ratings was 5.62 with a range of scores extending from 14 (perfect score) to -9 (-14 being the lowest possible score).

b. Classroom observation. The next process in determining the validity of the Inventory involved the writer's observation of the classroom behavior of each of the 100 teachers relative to his relationship with pupils. As an aid in this observation, use was made of Baxter's *Rating Scale of the Teacher's Personal Effectiveness* (1), modified so as to meet more adequately the needs of the present study. Baxter's Scale was devised with the principal object of evaluating a teacher's personal effectiveness in the classroom by considering the pupils' responses to his conduct. This scale was modified for use in the present investigation by (a) omitting the responses of the pupils, (b) omitting two of Baxter's items which were somewhat irrelevant to the present study, (c) adding an item relating to the teacher's disciplinary ability, and (d) slightly altering the wording of Baxter's items in several places. It was believed that such modification would facilitate the evaluation of each teacher's personal reactions toward the pupils.

The writer spent from twenty minutes to one hour in the classroom of each teacher to observe his effectiveness in the following:

- (1) Maintaining discipline.
- (2) Creating a friendly classroom atmosphere.
- (3) Establishing a feeling of security.
- (4) Exerting a stabilizing influence.
- (5) Developing pupil self-reliance.

The significance of the first two points above in teacher-pupil relationships is evident. Also important in the personal interaction between teacher and pupil is the feeling of security that a teacher is able to establish in his pupils. He may be encouraging and able to inspire confidence, or he may be intolerant of mistakes and definitely lacking in appreciation of the pupils' efforts.

Effectiveness in "exerting a stabilizing influence" essentially depends upon the teacher's own personality adjustment. The importance of this problem in teacher-pupil relationship cannot be over-emphasized.

A teacher's skill in developing pupil self-reliance is an important factor in his personal relationships with pupils. On the one hand, he may encourage pupil decision; on the other, he may demand rigid conformity.

The rating of each teacher relative to the above points provided a numerical estimate of his personal effectiveness with the pupils which served as one criterion in establishing the validity of the Inventory. The mean of the classroom observation scores was 3.60, and the range of scores extended from 10 to -9 (possible range from 10 to -10).

c. Attitudes of pupils. The third phase in determining the validity of the Inventory involved procuring the reactions of pupils to each of the 100 teachers. Of the three validating criteria, that considered of most importance was the attitudes of pupils toward their teachers. Who is in a better position to decide what a pupil thinks about his teacher than the pupil himself who is with him so much of each school day? Or, isn't what a pupil thinks about the teacher of much importance? A positive answer to this question repre-

sents a basic presupposition of the present study. Even though children are immature, when gauged by the standards of adults, and even though their opinions of teachers may change considerably as they grow older, it would seem that it is their *present* attitude, infantile as it may seem, that should be of greatest significance to the educator.

The measuring device constructed to procure an estimation of the pupils' attitudes toward the teacher took the form of a 50-item questionnaire entitled *My Teacher*.² The items were simple questions adapted to the mental maturity and experience of fourth-, fifth-, and sixth-grade children, that could be answered simply by underlining "Yes," "No," or "?."

The questions were formulated by the writer on the basis of ideas gleaned from the literature on teacher-pupil relationships. The content of the questions related to the teacher's disposition, his treatment of pupils, his teaching, his interest in children's activities, his status in the pupils' esteem, his sense of humor, and a number of other topics within the realm of teacher-pupil relationship. The list of questions was made long enough to exhaust the source material for item content, yet short enough that completion of the questionnaire would not tax the interest of a pupil in the intermediate grades. The following questions are cited as examples:

4. Does this teacher scold the pupils a lot? Yes No ?
6. Does this teacher explain the school work so that you can understand it? Yes No ?
11. Is this teacher often "bossy"? Yes No ?

²A copy of *My Teacher* may be found in Appendix D in the writer's Ph.D. dissertation (10) on file in the University of Minnesota Library.

38. Does this teacher give you a chance to ask questions? Yes No ?
 49. Do you like this teacher? Yes No ?

Space was provided on the questionnaire for the pupil to comment on (a) Why I Like This Teacher, and (b) Why I Don't Like This Teacher. The pupil was also asked to indicate the grade or mark which he usually received from this teacher.

As a safeguard against disturbing morale, classroom observation on the part of the writer was completed before pupil reactions were obtained.

In procuring these reactions, the general plan was to administer the questionnaire, in person, to groups of about 25 fourth-, fifth-, or sixth-grade pupils. As these grades were departmentalized, in whole or in part, in two of the three school systems used in this phase of the study, the majority of the pupils had classroom contact with several teachers during the school day. For practical reasons, in some cases a group of pupils rated only one teacher, whereas in other cases, a group rated two, or possibly three, teachers. However, only one teacher was rated during any one administration of the questionnaire. A condition that had to be met was that any teacher being rated by a group must have teaching contact, that school year, with the pupils in the group.

All the data for the investigation were gathered during the second half of the school year, at which time it was felt that teachers and pupils had had ample opportunity for mutual acquaintance.

To make conditions as uniform as possible, pupils were not asked to rate their home-room teachers. There is reason to believe that a home-room teacher occupies a somewhat different status in the thoughts and feeling of her home-room pupils from that of a teacher not serving in this capacity.

Each administration of the questionnaire (there were 100 separate administrations in all) required from fifteen to twenty minutes. Groups ranged in size from twenty to forty pupils, usually numbering about twenty-five. In some buildings it was the wish of the administrator that the writer go from room to room and use the whole class for each separate rating. In others it was considered desirable that the rating be done in one room. In the latter situation twenty-five pupils were selected at random from one of the intermediate grades by the principal

and directed to go to the room where the rating was being done.

The teacher who was being rated was never present during the administration of the questionnaire. In fact, no one was present other than the pupils doing the rating and the writer.

The principal made the following points clear to each group of pupils: (a) No one, including the teachers and the principal, would ever know how they (the individual pupils) answered the questions, for no one, except the writer, would ever see the papers; (b) the way the questions were answered would have no effect whatever upon their grades or school standing. In the interest of morale, the writer had also requested each principal to ask the pupils not to discuss the questions afterwards with teachers or other pupils. After these remarks, the principal left the room. The writer then stressed the desirability of honest answers, and re-emphasized the fact that he would be the only one that would ever see the papers, and that even he would not know how any individual pupil answered the questions as *no names were to be signed*. The writer then read the questions over with the pupils, making explanatory comments, and answering any questions that were asked.

The mean of the distribution of scores on the pupils' questionnaires was 20.0, the score range extending from 41 to -17 (possible range from 49 to -49).

d. *The reliability of the ratings.* The reliability of ratings is usually determined by correlating the ratings made by two comparable judges. In the present study, the use of this procedure did not appear feasible or practicable, particularly in connection with the ratings made by the principals and the writer. The method presupposes use of the same scale by both judges, whereas the rating scale used by the principals was different from the one used by the writer. Different scales were purposely employed. It was believed that a more extended and diversified list of traits to be rated, made possible by two *different* scales, would contribute definitely toward the validity of the ratings. Even if the same scale had been used by both principal and writer, correlation of the two ratings in deter-

mining the reliability of the instrument would have been questionable procedure in that the two judges would not have been "comparable" because of differences in background, mental set, and knowledge of the teacher being rated. The same problem would have arisen in any attempt to employ a third party to do the rating in the capacity of a second judge.

The reliability of the ratings made by the principals and the writer could have been determined by correlating two ratings made by the same individual at different times. It has been contended, however, that two ratings by the same individual are no better than one (18, p. 93). This fact coupled with the writer's hesitancy to impose upon the good will of school officials participating in the study led to the rejection of such a procedure.

(1) *Use of split-half technique.* The method used to determine the reliability of the rating scales was the split-half technique of correlating the odd items against the even items. Use of this method with rating scales is not very satisfactory because of the small number of items, the heterogeneity of their content, and the probable operation of the halo effect. The halo effect would tend to raise the coefficient of reliability, while at the same time lowering the validity of the scale. Regardless of its disadvantages, the split-half method was employed, and use was made of the Spearman-Brown formula to estimate the reliability of the whole scale from the calculated reliability of the half.

(2) *Reliability of the scale used by the principal.* By use of the split-half technique and the Spearman-Brown formula, the reliability coefficient of .87 was obtained for the principal's scale. Partly

responsible for this result, undoubtedly, were the halo effect and the so-called "logical error"—the tendency for a judge to give similar ratings in traits that seem logically related in his own thinking.

(3) *Reliability of the scale used by the writer.* A reliability coefficient of .92 for the scale used by the writer in classroom observation was obtained by means of the split-half method and the Spearman-Brown formula. The high value probably is due in part to such influences as the halo effect and the logical error.

(4) *Reliability of the questionnaire used by the pupils.* The coefficient of reliability was determined for the pupils' questionnaire, *My Teacher*, by correlating the odd items with the even items from 200 papers. These 200 papers were chosen by selecting, at random, two papers from each of the 100 piles of questionnaires relating to each of the 100 teachers. The result was found to be .94 after the Spearman-Brown formula was employed. It would seem that the halo effect and logical error would be especially potent in their influence upon ratings made by elementary school children. Too, the failure of the split-half method to take account of the fluctuational changes in individuals from day to day results in an estimate of reliability that is probably too high from the standpoint of practical testing.

Inasmuch as from twenty to forty pupils rated each teacher it was possible to compute the reliability of the instrument when used by more than one rater. Twenty papers chosen at random from each of the 100 piles (one pile for each of the 100 teachers) were divided into two piles of ten papers each. After finding the mean score for each pile of ten papers, the results were used in correlat-

ing the mean of ten ratings with the mean of the other ten ratings for each of the 100 teachers. The correlation coefficient was found to be .83. This value represents the reliability of the instrument for ten raters. Substituting 2.5 in the Spearman-Brown formula, the reliability coefficient, .93, was obtained as the estimate of reliability of the instrument for twenty-five raters. The value for twenty-five raters was determined inasmuch as the majority of teachers were rated by groups of twenty-five pupils each.

It would seem that the value, .93, obtained for twenty-five raters, is of greater significance in the practical use of the questionnaire than the value, .94, which was the estimated reliability of the instrument as such when used by one rater. The instrument would ordinarily be used by a class of pupils rather than by a single pupil. Furthermore, the value of the reliability coefficient determined on either ten or twenty-five raters would appear to be closer to the true reliability sought when one considers the errors in rating previously discussed in connection with the split-half technique.

4. Comparisons Among Ratings and Inventory Scores

a. *The distribution of Inventory scores for the unselected group of teachers.* It is of interest to consider the distribution of Inventory scores for the unselected group of teachers in relation to the distribution of scores for the two

TABLE 10
RESULTS OBTAINED IN THE COMPARISON OF
INVENTORY SCORES FOR THE SUPERIOR,
UNSELECTED, AND INFERIOR GROUPS
OF TEACHERS

| Statistic | Superior Group | Unselected Group | Inferior Group |
|------------|-------------------|---------------------|-------------------|
| N | 100 | 100 | 100 |
| Range | 324-(-99) | 254-(-176) | 213-(-297) |
| M | 131.0 | 77.6 | -32.0 |
| S.D. | 102.91 | 103.49 | 112.46 |
| σ_M | 10.19 | 10.35 | 11.18 |
| Mdn | 130.0 | 85.5 | -41.5 |

original groups of teachers. Table 10 indicates the range of scores, mean, standard deviation, and median for each of the three distributions. The data are based upon the Inventory scores obtained by the original scoring process.

As one would expect, both measures of central tendency (mean and median) for the unselected group lie between the central tendency values obtained for the superior and inferior groups. However, the obtained mean (77.6) for the unselected group is somewhat larger than the average (49.5) of the means for the two extreme groups. The same holds true for the obtained median (85.5) as compared with the average (44.2) of the other two medians. These differences may be due to one or more of such factors as (a) the limitation of the unselected group to teachers of the fourth, fifth, and sixth grades (seventh- and eighth-grade teachers in original samples were lowest); (b) the possibility that, of the two original groups of teachers, the inferior group was farther below the

TABLE 11
VALUE AND SIGNIFICANCE OF g_1 AND g_2 FOR THE DISTRIBUTION OF INVENTORY SCORES MADE
BY THE UNSELECTED GROUP OF TEACHERS

| Group | g_1 | t_{g_1} | Probability of t_{g_1} | g_2 | t_{g_2} | Probability of t_{g_2} |
|------------|---------|-----------|-----------------------------|---------|-----------|-----------------------------|
| Unselected | -.37845 | 1.576 | >.05 | -.51469 | 1.076 | >.05 |

parametric mean of an unselected population than the superior group was above it (weak teachers possibly more easily identified by principals than strong teachers); (c) the possibility that the unselected group was actually composed of better-than-average teachers; (d) chance errors.

It should also be observed that the range of the scores for the unselected group is from 254 to -176, whereas the total range for the extreme groups was from 324 to -297. The mean of the scores obtained by the simplified scoring system is 88.4, and the range extends from 120 to 38.

The distribution of Inventory scores for the unselected group was tested for normality as were the two original distributions. The statistics g_1 and g_2 were again employed. Table 11 presents the results obtained. The negative value of g_1 indicates an excess of scores larger than the mean or a mean smaller than the median. The value of g_1 does not differ significantly from zero, however, and asymmetry is not indicated. Although the negative value of g_2 indicates a tendency toward a flat-topped distribution curve, the value of g_2 is too small to

be significant. As with each of the two original distributions, there is little evidence of departure from normality in the distribution of scores for the unselected group of teachers.

b. *Correlations between Inventory scores and rating scale scores.* Validity coefficients were found by computing correlations between scores on the revised Inventory and scores on each of the teacher rating devices. Table 12 indicates the coefficients obtained in comparing the Inventory results (both original and simplified scoring) with results of the ratings considered separately and in combination. Results obtained in correlating the ratings with each other are also presented. The single validating criterion showing the highest correlation (.486) with the Inventory is the writer's observation of the teacher's classroom behavior, whereas the criterion showing the lowest correlation (.434) is the principals' ratings. Between these values is found the correlation (.452) of the Inventory with the attitudes of the pupils. The fact that, of the three criteria, the

TABLE 12
CORRELATIONS OBTAINED BETWEEN INVENTORY SCORES (ORIGINAL AND SIMPLIFIED SCORING) AND RATINGS CONSIDERED SEPARATELY AND IN COMBINATION, TOGETHER WITH CORRELATIONS OBTAINED BETWEEN THE RATINGS, FOR THE UNSELECTED GROUP OF 100 TEACHERS

| | Original Scoring | | Simplified Scoring | |
|--|------------------|------------|--------------------|------------|
| | r^* | σ_r | r^* | σ_r |
| Inventory and Principals' Ratings..... | .434 | .082 | .445 | .081 |
| Inventory and Classroom Observations..... | .486 | .077 | .485 | .077 |
| Inventory and Pupils' Attitudes..... | .452 | .080 | .456 | .080 |
| Inventory and Three Validating Criteria Combined..... | .594 | .065 | .596 | .065 |
| | .595** | .065 | | |
| Inventory and Combination of Classroom Observations and Principals' Ratings..... | .536 | .072 | .544 | .071 |
| Inventory and Combination of Classroom Observations and Pupils' Attitudes..... | .576 | .067 | .579 | .067 |
| Inventory and Combination of Principals' Ratings and Pupils' Attitudes..... | .534 | .072 | .543 | .071 |
| Classroom Observations and Principals' Ratings.... | .480 | .077 | | |
| Classroom Observations and Pupils' Attitudes..... | .326 | .090 | | |
| Principals' Ratings and Pupils' Attitudes..... | .389 | .085 | | |

* These obtained r 's are all statistically significant at the 1 per cent level (8, p. 324).

** Multiple correlation coefficient.

principals' ratings show the lowest correlation is rather interesting, inasmuch as the judgments of principals formed the criterion on which the selection of the original standardization groups was made.

It was believed that the three validating criteria combined, or two of them in the several possible combinations, would prove of greater value in establishing the validity of the Inventory than any one taken alone. It seemed reasonable to suppose that the combined judgment of several persons or groups relative to a teacher's behavior is of greater worth than the judgment of any one single agency. To make a combination of measures possible, all raw scores were converted into standard *T*-scores.

The correlation between the Inventory scores and the three validating criteria combined was found to be .594 (.596 with simplified scoring). When several criteria are thus considered together, extremes in one or two of them are counterbalanced by more moderate ratings in those remaining. A "tempering" or "leveling" process takes place which, apparently, tends to bring the resulting score into closer relationship with the score on the Inventory.

A multiple correlation of .595 between the Inventory scores and the three criteria was also obtained.

Of the several combinations of two validating criteria, that of the classroom observations and pupils' attitudes correlates highest (.576) with the Inventory. When compared with each other, however, these two criteria show the least agreement (r equals .326). In closest agreement are the classroom observations and the principals' ratings (r equals .480). The fact that the ratings of the principals and those of the writer

agree more closely with each other than either of these agrees with the reactions of the pupils is of interest. Such findings suggest possible differences in standards of evaluation by which adults and children judge the behavior of teachers. Support is also given the view that pupil rating is an important and unique contribution to an adequate evaluation of teacher behavior.

Table 12 also presents the correlation coefficients computed when using the Inventory scores obtained by use of the simplified scoring procedure. It is interesting to note that, in each case except one (comparison between Inventory and classroom observations), these correlations are slightly higher than those obtained when using the original Inventory scores. These slight differences, though quite consistent, are inconsequential, however, in comparison with the close agreement between the two sets of values.

5. *An Additional Check on the Validity of the Items*

An additional check was made on the validity of the items in the revised instrument. Out of the distribution of the 100 teachers in the unselected group the item response patterns of the twenty with the highest scores and the response patterns of the twenty with the lowest scores were compared, respectively, with the patterns of item response characterizing the original groups of superior and inferior teachers. Chi-square was the statistic used as the measure of comparison.

It seemed reasonable to suppose that, if a particular item were valid, it would call forth similar patterns of response from the two superior groups of teachers and similar patterns from the two in-

ferior groups. Of the 328 (2 x 164) comparisons, 270, or 82 per cent of them had chi-square values less than 9.488, a value which is at the 5 per cent level of significance. A chi-square value less than 9.488, therefore, indicated a high probability that differences found between the two groups, in response to a particular item, could be chance differences. Al-

agreement or disagreement but reveal discrepancies in the *degree* of reaction. The group of twenty teachers, at each extreme, generally tended to exhibit a more intensive attitude relative to the opinion statements.

B. DETERMINING THE RELIABILITY OF THE INVENTORY

The split-half method was used in determining the reliability of the *Teacher-*

TABLE 13
DISTRIBUTION OF RESPONSES OF THE SUPERIOR AND INFERIOR GROUPS OF TEACHERS AND OF THE HIGHEST 20 AND LOWEST 20 TEACHERS IN THE UNSELECTED GROUP TO ITEM 11

| 11. Pupils are naturally lazy when it comes to doing school work. | | | | | |
|---|----------------|-------|-----------|----------|-------------------|
| | Strongly Agree | Agree | Uncertain | Disagree | Strongly Disagree |
| Group | 1 | 2 | 3 | 4 | 5 |
| Superior | 0 | 7 | 11 | 73 | 9 (N = 100) |
| Inferior | 3 | 17 | 14 | 63 | 3 (N = 100) |
| Highest 20 | 1 | 2 | 0 | 9 | 8 (N = 20) |
| Lowest 20 | 1 | 5 | 2 | 10 | 2 (N = 20) |

though a higher proportion than 82 per cent would have been desirable, this result does lend further support to the validity of the items in the Inventory.

It is of interest to analyze further, by inspection, the response patterns of an item that seems to be representative of those cases where the chi-square value was greater than 9.488. The response patterns for Item 11 are indicated in Table 13. The chi-square value of 20.618 was found, for Item 11, on comparing the response patterns of the 100 superior teachers with those of the highest twenty in the unselected group. Contributing to the high value of chi-square are the differences between the two groups relative to responses "4" and "5." The great majority of both groups of superior teachers disagree with the statement. Yet, a much larger proportion (40 per cent) of the twenty teachers "strongly disagree" than of the 100 teachers, where only 9 per cent "strongly disagree." This relatively high proportion (40 per cent) undoubtedly accounts for the small representation found under responses "3" and "4" of the group of twenty teachers. As is true with the other items showing high chi-square values, the response patterns indicate validity of the item relative to general

Pupil Inventory. Correlating the scores on the odd-numbered items with those on the even-numbered items, a correlation coefficient of .833 was obtained. Use of the Spearman-Brown prophecy formula resulted in a reliability coefficient of .909 for the entire instrument.

The reliability of the Inventory was also determined, by means of the split-half method, on the basis of the scores obtained by the simplified scoring procedure. Under these conditions the odd-even correlation was .794. Determined by use of the Spearman-Brown formula, the reliability coefficient for the entire instrument was found to be .885. The standard error of measurement, determined on the basis of the simplified scoring system, was 6.25. May it be mentioned again at this point that the simplified scoring procedure yielded results which correlated .973 with the original scores.

IV. SUMMARY AND CONCLUSIONS

1. The major purpose of this investigation was to construct a measuring instrument which would gauge the attitudes of teachers toward pupils and serve to differentiate those teachers who get along well with children from those who do not. The instrument, known as the *Teacher-Pupil Inventory*,¹ in its revised form, consists of 164 statements of opinion relating to pupils and children and the part the teacher should play in their educational development. Administration time is from twenty to thirty minutes.

For purposes of standardization and scoring, the original instrument, consisting of 756 items (2 forms with 378 items each), was administered to 100 teachers in each of two groups, one at each extreme of the teacher-pupil relationship continuum, and comparison then made of the item responses. The statistic, chi-square, served as the objective measure of comparison. Those items not differentiating adequately were rejected. A system of scoring was developed on the basis of the responses of the superior and inferior groups of teachers.

2. Significant aspects of the present investigation included the use of empirical procedures of test construction and the use of pupil reaction as a basic validating criterion. The validity of individual items in the instrument was established by the use of the chi-square technique in the comparison of the response patterns provided by the two groups of teachers at the extremes of

teacher-pupil relationship. The validity of the Inventory as a whole was determined by administering it to a new sample of 100 teachers from the fourth-, fifth-, and sixth-grade levels, *unselected* as to teacher-pupil relationship, on whom ratings were then obtained from each of three sources: (a) the principal, (b) the writer (from classroom observation), and (c) the pupils.

When correlated with Inventory scores, the ratings of the principal, the writer, and the pupils gave validity coefficients of .434, .486, and .452, respectively. After combining the three ratings, the combined result correlated .594 (.596 with simplified scoring) with the Inventory scores. A multiple correlation of .595 between the Inventory and the three criteria was also obtained. These correlations were all significant at the 1 per cent level.

The means of the Inventory scores of the superior, unselected, and inferior groups of teachers were, respectively, 131.0, 77.6, and -32.0. The difference between the means of the superior and inferior groups yielded a critical ratio of 10.8. By statistical test (use of g_1 and g_2), none of the three distributions of Inventory scores was found to depart significantly from normality.

3. The reliability of the *Teacher-Pupil Inventory* was found to be .909 as determined by means of the split-half method and the Spearman-Brown prophecy formula. On the basis of scores obtained by the simplified scoring system, the reliability coefficient was .885.

This investigation has shown that teachers' attitudes toward pupils and

¹ The two original forms and the revised form of the *Teacher-Pupil Inventory* may be found in Appendices E, F, and G, respectively, in the writer's Ph.D. dissertation (10) on file in the University of Minnesota Library. Plans are being made to publish the Inventory in the near future.

their behavior are related to teacher-pupil rapport in the classroom. It has shown further that these attitudes can be measured with as high a validity (.595) as can academic aptitude.

These findings are of significance in teacher training, selection, and evaluation. It is true that the validity of the Inventory was determined under favorable conditions of test administration. Further research² is needed to demon-

² Research is under way which should lead to a more refined instrument and a greater knowledge of its measurement possibilities.

strate the validity of the instrument for purposes of selection and evaluation of teachers under actual "operating conditions" of administration.

The day may not be far distant when teacher-training institutions will be able to take account of a teacher's personal effectiveness in the classroom in as objective a manner as is done at present with teaching techniques and knowledge of subject matter. It is hoped that such an instrument as the *Teacher-Pupil Inventory* will contribute, in a small but significant way, to this end.

BIBLIOGRAPHY

1. BAXTER, BERNICE. *Teacher-pupil relationships*. New York: Macmillan, 1942.
2. BRYAN, R. C. Why student reactions to teachers should be evaluated. *Educ. Adm. & Superv.*, 1941, 17, 590-603.
3. DROBA, D. D. The nature of attitude. *J. soc. Psychol.*, 1933, 4, 444-463.
4. DUNLAP, J. W., & DROLL, A. Observations on the methodology in attitude scales. *J. soc. Psychol.*, 1939, 10, 475-487.
5. FENTON, NORMAN. *Mental Hygiene in School Practice*. Stanford, Calif.: Stanford Univ. Press, 1943.
6. FERGUSON, L. W. The requirements of an adequate attitude scale. *Psychol. Bull.*, 1939, 36, 665-673.
7. GARRETT, HENRY E. *Statistics in Psychology and Education*. New York: Longmans, Green, 1939.
8. GUILFORD, J. P. *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill, 1942.
9. KELLEY, IDA, & PERKINS, KEITH. An investigation of teachers' knowledge of and attitudes toward child and adolescent behavior in everyday school situations. *Further Studies in Attitudes*, Series IV, Stud. in Higher Educ., 42, Purdue University, 1941.
10. LEEDS, CARROLL H. *The Construction and Differential Value of a Scale for Determining Teacher-Pupil Attitudes*. Unpublished Ph.D. thesis on file in the University of Minnesota Library, 1946.
11. LIKERT, R. A technique for the measurement of attitudes. *Arch. Psychol.*, 1932, 22, No. 140.
12. PRESCOTT, D. A. *Emotion and the educative process*. Washington, D.C.: American Council on Education, 1938.
13. REMMERS, H. H. Generalized attitude scales. *J. soc. Psychol.*, 1934, 5, 298-312.
14. ROSANDER, A. C. Attitude scale based upon behavior situations. *J. soc. Psychol.*, 1937, 8, 3-15.
15. SNEDECOR, GEORGE W. *Statistical methods*. Ames, Iowa: Collegiate Press, 1938.
16. STOGDILL, R. M. *Measurement of attitudes toward children*. Ph.D. thesis. Columbus, Ohio: Ohio State Univ., 1934.
17. STRONG, E. K., JR. *Vocational Interests of Men and Women*. Stanford, Calif.: Stanford Univ. Press, 1943. Chapters 23 and 24.
18. SYMONDS, P. M. *Diagnosing personality and conduct*. New York: Appleton-Century, 1931. Chapters 3 and 6.
19. TENENBAUM, S. Test to measure a child's attitude toward school teachers and classmates. *Educ. Adm. & Superv.*, 1940, 26, 176-188.
20. THURSTONE, L. L., & CHAVE, E. J. *The measurement of attitude*. Chicago: Univ. of Chicago Press, 1929.
21. TIEDEMAN, S. C. Study of pupil-teacher relationship. *J. educ. Res.*, 1942, 35, 657-664.
22. TSCHICHTELIN, M. A. & OTHERS. Measuring the attitudes of elementary school children toward their teachers. *J. educ. Psychol.*, 1940, 31, 195-203.
23. WICKMAN, E. K. *Children's behavior and teachers' attitudes*. New York: Commonwealth Fund, 1928.